Chapter 2

General Concepts

2-1. General

These concepts define and describe the components which constitute the Command Data Model (CDM) environment. A basic, Corps-wide understanding of these concepts is necessary to ensure consistency when referring to the Command Data Model and Dictionary. The importance of a common structure and definition of data cannot be overemphasized The Corps must share its data, must communicate its data, and must understand its data.

2-2. Information and Data

The Corps defines information as a representation of facts, concepts, and assumptions that are derived from data for use in performing the various Corps' missions. Data is defined as meaningful facts about persons, places, things, concepts, events, and activities in a defined format and structure from which information may be derived. Data is considered a Command resource that is shared among many users as appropriate. Data is manipulated and processed to create information from which business decisions can be made. Identifying the role that the data plays in deter-mining the Corps' business rules is an essential part of identifying reliable data.

2-3. Data Objects

Data objects are entities, data elements, relationships, and data models and their characteristics when structured, combined, and analyzed, provide an understanding of the business process.

2-4. Entities

a. Entities are the persons, places, things, concepts events, or activities about which we want to collect data. The data object, entity, is a logical place to store the characteristics of the real world objects. Analysts show the relationships between entities and the rules affecting their relationship graphically. The entities and their relationships are the building blocks used to depict, or model, the data needed to support a business process.

b. The USACE has categorized data into areas or classes of information. See Appendix C “USACE Data Classes," for an explanation and a listing of USACE Data Classes.

2-5. Data Elements

a. Data elements are uniquely defined and identifiable characteristics of the person, place, or thing defined by an entity. When values are given to these data elements, the values represent facts about specific instances of an entity important to the business. Data elements are the smallest unit of meaningful data.

b. The definition of a data element includes a description of the characteristic that it represents, such as CONTRACT-START-DATE, the date that the contractor may begin work under a specific contract. The definition also includes the Domain which is the set of values that a data element may assume, usually defined by the data element's length and

format, or by a specific list of values. (Refer to Appendix D for lists of Common and Standard Domains.)

c. To provide continuity from current (legacy) data systems to the Command Data Model (CDM), it is helpful to see the correlation between the Legacy System Data Elements and the Command Data Elements. This is further discussed in Appendix E, “Legacy System to CDM Correlation."

2-6. Naming Conventions

When dealing with data objects (e.g., entities and data elements), it is important to understand and agree on the object's name and definition. As a system is developed, the structure and definition of all data should be stored in the Data Encyclopedia or Dictionary. The Data Encyclopedia is a valuable tool for organizing, tracking, and controlling the vast amounts of data used by the Corps. In order to effectively use this tool, naming conventions have been established that help users to quickly locate data objects and identify duplicate data objects. The Corps of Engineers “Data Element Defining and Naming Conventions" appear in Appendix F of this document.

2-7. USACE Data Encyclopedia

a. The USACE Data Encyclopedia addresses the Corps' needs by providing users with an information sharing tool, a repository of metadata (information about data), and an automated cross reference resource tool (for cross referencing activity and data models) used in the information resource methodology adopted by the Corps.

(1) The methodology for Information Systems development includes:

(a) Requirements Analysis Planning

(b) Conceptual Design

(c) Iterative Application Development

(2) Throughout the process of this methodology, the data administrative functions will occur as follows:

(a) In concert with the modeling of the business, the information needed by the various business activities is defined. This information includes the Inputs, Controls, Outputs, and Mechanisms (ICOMs) used by the business people in their activities.

(b) Entities are identified which will contain the fundamental data elements represented in the inputs, controls, outputs and mechanisms as well as other data that needs to be kept; then an Entity-Relationship (E-R) Data Model is developed.

(c) The entities are defined.

(d) The entities key data elements are determined and the relationships among entities reviewed.

(e) The non-key data elements that describe the entity are identified.

(f) All data elements are defined.

b. The Data Encyclopedia will facilitate each stage of this process. It will: Improve

system documentation; provide accurate, timely metadata during system development; provide extensive cross references of corporate data; help coordinate the documentation needs of all users; and assist in enforcing standards.

c. The inquiries and listings available from the Data Encyclopedia will enable management and users to spot redundant data. Redundant data are likely to cause inconsistencies in the information derived from it. The same piece of data may have different formats or many versions, resulting in inconsistent, or even inaccurate business decisions.

d. The benefits provided by the USACE Data Encyclopedia during system development include: The minimization of data redundancy; ease in developing and maintaining current documentation (therefore, minimizing change impact); and extensive cross references and multiple views of existing Corps data sources.

2-8. IDEF Modeling Technique

The U.S. Air Force's Integrated Computer Aided Manufacturing (ICAM) invented a technique required on many of its projects which evolved into the Integrated Computer Aided Manufacturing Definition, ICAM Definition (IDEF) Modeling Technique. The IDEF Modeling Techniques have been adopted by the Corps of Engineers to be used during system development. The Corps uses the IDEF Modeling Techniques in the development of activity and data models which graphically and descriptively depict the Corps' business processes. The IDEF Modeling Techniques provide: An effective means for displaying the current (AS-IS) and the planned (TO-BE) environments; a means to accomplish interaction and consensus between user groups as part of system development; consistency, which facilitates data integration and sharing; and the use of documentation and audit trails during system conceptual design.

2-9. Command Data Dictionary CDD

a. The Command Data Dictionary is a printed glossary of data terms. Its information can also be retrieved online from the Encyclopedia system. If Webster's dictionary is essential to clear communication in the English language, then data structures and definitions in the Encyclopedia are essential to clear communication within systems design, development, and use.

b. The Data Encyclopedia provides a mechanism to define and use information about data. It is a central repository of this information about data called metadata. Consistent with the increase in the efficiency of methods used to collect, compute, and disseminate mission data is the need to understand the characteristics and relationships of the data itself. The Encyclopedia provides the technology to record, analyze, and share this. The data dictionary is a report of the metadata about the Corps' data. The Corps' CDM and CDD are based on the IDEA Modeling Techniques.

2-10. Command Data Model (CDM)

a. The Command Data Model is a graphical representation of the structure of the USACE corporate data. It contains the

entities, relationships, and data elements of the Corps' data. Like the Command Data Dictionary, it is a representation of the metadata that describes the important shared data of the Corps. This metadata is stored in the Encyclopedia. For the CDM, it is a graphical representation. It serves as a communication tool between functional and technical users.

b. Because it is graphical, it best shows the structure and, of course, the complexity of the totality of the Corps' data that has been defined as of a particular point in time. But, because it is graphical, the full definitions cannot be shown on it without destroying its structural message. Thus, the Command Data Dictionary report or the many Encyclopedia search and retrieval functions are used to present this detail.

c. The Command Data Model and Command Data Dictionary (CDM/CDD) are the base from which shared information will evolve in the Corps. The model flow (in Figure 2-10) shows how the CDM/CDD serves to:

(1) Provide a starter model for a new project.

(2) Integrate project final data models after they have been validated and have passed the In-Process Reviews (IPRs) and alpha tests.

(3) Interface legacy systems to the CDM.

(4) Provide a physical database design from a conceptual project final model.

NOTE: Sequencing numbers denote normal development cycle.

2-11. Extracting a Data Model View

a. The Encyclopedia has the capability to extract a subset from a model. This feature is invoked by first creating a Data Model Header for the new model, then following the instructions for extracting a data model which is accessed under the Encyclopedia menu of "Other Data Model Functions."

b. The new data model is populated by selecting the relationships that are going into it which indirectly selects the entities of choice. If the entities selected are "Registered" or "Approved," they can be only borrowed or shared because they represent Corps-wide entities, approved and controlled by the DACC. A new project can create "Candidate" entities and relationships. Relationships can be deleted only if they do not affect the key structure of the CDM.

2-12. Fundamental and Derived Data

Fundamental data is the lowest level, basic piece of information or fact. The data model and dictionary deal with basic or fundamental data. Derived data is information prepared by algorithms, conditional extractions, or formulas, such as "the sum of costs from 7/89 thru 10/89" or "inflation adjusted project estimate." The latter is derived using the data elements PROJECT-ESTIMATE and INFLATION-RATE.

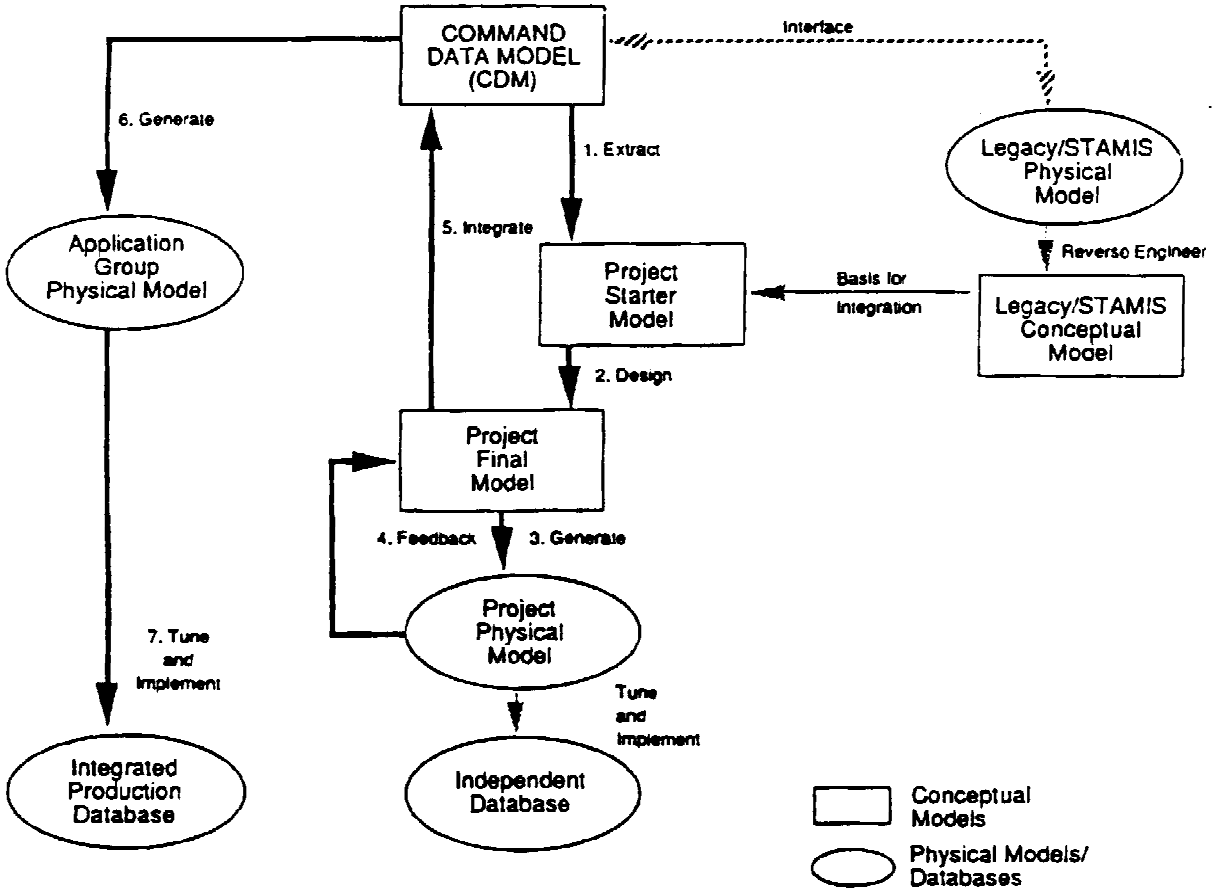

Figure 2-1: USACE Encyclopedia Data Administration & Database Administration Models, Operations, and Interfaces

2-13. Conceptual Data Models vs Physical Data Models

a. The Conceptual Data Model (or Command Data Model in the Corps) is the normalized basic structure of the business rules, entities, and data elements that describe the things the Corps needs to operate. A conceptual model deals with the fundamentals, following specific rules for accuracy and communication The conceptual data model is a neutral view of the data. For example, an “inflation adjusted cost estimate" has the same meaning whether it is seen from an accounting viewpoint' a project manager's viewpoint' or the viewpoint of virtually any function within the Corps.

b. When a subset of the CDM (Command Data Model) is formed into a specific database to support one or more applications, the may be adaptations made for

performance or security reasons. The result is called an physical model. The physical model may contain derived values, may combine or split entities into records that suit the performance requirements better, or may even repeat data elements in different records.

c. The following CDD and CDM relate to physical models in that:

(1) They provide the conceptual metadata which includes the fundamental data and definitions, the building blocks, of the physical model.

(2) Many (or most) entities will become a table in the physical model without change.

(3) The fields of the physical model usually correspond to data elements in the conceptual model although sometimes there may be derived fields in the physical model that combine multiple conceptual data elements algorithmically (such as: summation, ratio, or complex formula).

(4) The domain attributes specified in the dictionary are recommended for use in physical models throughout the Corps and provide the rules for input editing.

2-14. How are the CDM and the CDD Used in Routine MSC Work?

a. The IDEF Modeling Techniques and the continuing evolution of the Command Data Model and the Command Data Dictionary will assist the MSCs in these ways:

(1) Help the field Information Management (IM) staffs develop better systems.

(2) Help information users throughout the Corps to better understand the data with which they are dealing.

(3) Facilitate data sharing and the interchange of data.

(4) Identify and promote the use of a common set of data structures that represent the way the Corps does business.

(5) Standardize data definitions and do- mains, including data type, size, and standard lists of values where appropriate (such as, organization code, project status code, etc.)

b. In the pursuit of happiness and better Information Systems throughout the field, you will find it useful to:

(1) Familiarize yourself with the overall structure of the CDM. Use it as a roadmap of the CDD.

(2) If you are beginning a new application development, see if you can extract a starter data model from the CDM in order to first make your job easier, and second, tie in to the structures that other Corps IM projects are using.

(3) Browse through the CDM and CDD to see if there are entities or data elements that have already been defined that you can use in your systems development.

(4) Check any system or database that you must interface with to see if any of the data that you will be sharing is already defined in the CDM/CDD. If not, contact Data Administration because you have just discovered data that needs to be shared. That data needs to be incorporated into the CDM/CDD so the two known users and any more that come along can be assured of consistency.

(5) If the CDM/CDD structures or definitions don't work for you, let Data Administration know. The evolution and quality of the CDM/CDD can only occur if you make it happen.

c. Later, after the next development phase of the Encyclopedia, we will include suggestions for the inclusion of current and legacy systems into the Encyclopedia.